DNA sequence, and recombinant preparation of group 4 major allergens from cereals

Background of the invention

The present invention relates to the provision of DNA sequences of group 4 major allergens from cereals (*Triticeae*). The invention also encompasses fragments, new combinations of partial sequences and point mutants having a hypoallergenic action. The recombinant DNA molecules and the derived polypeptides, fragments, new combinations of partial sequences and variants can be utilised for the therapy of pollen-allergic diseases. The proteins prepared by recombinant methods can be employed for *in vitro* and *in vivo* diagnosis of pollen allergies.

Type 1 allergies are of importance worldwide. Up to 20% of the population in industrialised countries suffer from complaints such as allergic rhinitis, conjunctivitis or bronchial asthma. These allergies are caused by allergens present in the air (aeroallergens) which are released by sources of various origin, such as plant pollen, mites, cats or dogs. Up to 40% of these type 1 allergy sufferers in turn exhibit specific IgE reactivity with grass pollen allergens, inter alia cereal pollen allergens (Freidhoff et al., 1986, J. Allergy Clin. Immunol. 78, 1190-2001). Of the cereal pollen allergens, the allergens of rye have particular importance.

The substances which trigger type 1 allergy are proteins, glycoproteins or polypeptides. After uptake via the mucous membranes, these allergens react with the IgE molecules bonded to the surface of mast cells in sensitised individuals. If two IgE molecules are crosslinked to one another by an allergen, this results in the release of mediators (for example histamine,

prostaglandins) and cytokines by the effector cell and thus in the corresponding clinical symptoms.

A distinction is made between major and minor allergens, depending on the relative frequency with which the individual allergen molecules react with the IgE antibodies of allergy sufferers.

The allergens from the pollen of various species from the family of the grasses (*Poaceae*) are divided into groups which are homologous amongst one another.
In particular, the molecules of major allergen group 4 have high immunological cross-reactivity with one another both with monoclonal murine antibodies and also with human IgE antibodies (Fahlbusch et al., 1993 Clin. Exp. Allergy 23:51-60; Leduc-Brodard et al., 1996, J. Allergy Clin. Immunol. 98:1065-1072; Su et al., 1996, J. Allergy Clin. Immunol. 97:210; Fahlbusch et al., 1998, Clin. Exp. Allergy 28:799-807; Gavrovic-Jankulovic et al., 2000, Invest. Allergol. Clin. Immunol. 10 (6):361-367; Stumvoll et al. 2002, Biol. Chem. 383:1383-1396; Grote et al., 2002, Biol. Chem. 383:1441-1445; Andersson and Lidholm, 2003, Int. Arch. Allergy Immunol. 130:87-107; Mari, 2003, Clin. Exp. Allergy, 33 (1):43-51).

A complete DNA sequence is hitherto not known for any of the group 4 major allergens.

From the group 4 allergen from *Dactylus glomerata*, it has hitherto only been possible for peptides to be obtained by enzymatic degradation and sequenced:
DIYNYMEPYVSK (SEQ ID NO 13),
VDPTDYFGNEQ (SEQ ID NO 14),
ARTAWVDSGAQLGELSY (SEQ ID NO 15)
and GVLFNIQYVNYWFAP (SEQ ID NO 16, Leduc-Brodard et al., 1996, J. Allergy Clin. Immunol. 98: 1065-1072).

Peptides have also been obtained from the group 4 allergen of sub-tropical Bermuda grass (*Cynodon dactylon*) by proteolysis and sequenced:
KTVKPLYIITP (SEQ ID NO 17),
KQVERDFLTSLTKDIPQLYLKS (SEQ ID NO 18),
TVKPLYIITPITAAMI (SEQ ID NO 19),
LRKYGTAADNVIDAKVVDAQGRLL (SEQ ID NO 20),
KWQTVAPALPDPNM (SEQ ID NO 21),
VTWIESVPYIPMGDK (SEQ ID NO 22),
GTVRDLLXRTSNIKAFGKY (SEQ ID NO 23),
TSNIKAFGKYKSDYVLEPIPKKS (SEQ ID NO 24),
YRDLDLGVNQVVG (SEQ ID NO 25),
SATPPTHRSGVLFNI (SEQ ID NO 26)
and AAAALPTQVTRDIYAFMTPYVSKNPRQAYVNYRDLD (SEQ ID NO 27, Liaw et al., 2001, Biochem. Biophys. Research Communication 280: 738-743).

For *Lolium perenne*, peptide fragments having the following sequences have been described for the basic group 4 allergen: FLEPVLGLIFPAGV (SEQ ID NO 28) and GLIEFPAGV (SEQ ID NO 29, Jaggi et al., 1989, Int. Arch. Allergy Appl. Immunol. 89: 342-348).

As the first sequence of a group 4 allergen, the still unpublished sequence of Phl p 4 from *Phleum pratense* (SEQ ID NO 11) has been elucidated by the inventors of the present patent application and described in International Application WO 04/000881.

Nothing is hitherto known on the sequences of the group 4 major allergens from cereals (*Triceae*).

The object on which the present invention was based therefore consisted in the provision of DNA sequences of group 4 major allergens from cereals, in

particular the allergen Sec c 4 from rye (*Secale cerale*) (SEQ ID NO 1, 3), Hor v 4 from barley (*Hordeum vulgare*) (SEQ ID NO 5) and Tri a 4 from wheat (*Triticum aestivum*) (SEQ ID NO 7, 9) and of corresponding recombinant DNA molecules on the basis of which the allergens can be expressed as protein and made available, as such or in modified form, for pharmacologically significant exploitation. The sequence of Phl p 4 (SEQ ID NO 11) was the starting point for the present invention.

List of sequences according to the invention

The DNA and protein sequences of the mature allergens in accordance with SEQ ID NO 1-10 are preceded by a signal sequence. The encoding region ends with the TGA or TAG stop codons in the DNA sequences.

- DNA sequence of Sec c 4. (a) Isoform Sec c 4.01 (SEQ ID NO 1), (b) isoform Sec c 4.02 (SEQ ID NO 3).
- Protein sequences (SEQ ID NO 2, 4) derived from the DNA sequences in accordance with SEQ ID NO 1 and 3.
- DNA sequence of Hor v 4 (SEQ ID NO 5).
- Protein sequence (SEQ ID NO 6) derived from the DNA sequence in accordance with SEQ ID NO 5.
- DNA sequence of Tri a 4. (a) Isoform Tri a 4.01 (SEQ ID NO 7), (b) isoform Tri a 4.02 (SEQ ID NO 9).
- Protein sequences (SEQ ID NO 8, 10) derived from the DNA sequences in accordance with SEQ ID NO 7 and 9.
- DNA sequence of Phl p 4 (SEQ ID NO 11), in accordance with SEQ ID NO 5 from WO 04/000881.
- Protein sequence of Phl p 4 (SEQ ID NO 12), in accordance with SEQ ID NO 6 from WO 04/000881.

Description of the invention

The present invention now provides for the first time DNA sequences of the cereal pollen major allergens Sec c 4, Hor v 4 and Tri a 4, in accordance with SEQ ID NO 1, 3, 5, 7, and 9.
The present invention therefore relates to DNA molecules selected from the nucleotide sequences in accordance with SEQ ID NO 1, 3, 5, 7, and 9.

The invention furthermore relates to sequences homologous to the DNA sequences according to the invention and corresponding DNA molecules of group 4 allergens from other *Poaceae*, such as, for example, *Lolium perenne, Dactylis glomerata, Poa pratensis, Cynodon dactylon* and *Holcus lanatus*, which, owing to the sequence homology that exists, hybridise with the DNA sequences according to the invention under stringent conditions, or have immunological cross-reactivity with respect to the allergens according to the invention.

The invention also encompasses fragments, new combinations of partial sequences and point mutants having a hypoallergenic action.

The invention therefore furthermore relates to corresponding partial sequences, a combination of partial sequences, or replacement, elimination or addition mutants which encode an immunomodulatory, T-cell-reactive fragment of a group 4 allergen from the *Poaceae*.

With knowledge of the DNA sequence of the naturally occurring allergens, it is now possible to prepare these allergens as recombinant proteins which can be used in the diagnosis and therapy of allergic diseases (Scheiner and Kraft, 1995, Allergy 50: 384-391).

A classical approach to effective therapeutic treatment of allergies is specific immunotherapy or hyposensitisation (Fiebig, 1995, Allergo J. 4 (6): 336-339, Bousquet et al., 1998, J. Allergy Clin. Immunol. 102 (4): 558-562). In this method, the patient is injected subcutaneously with natural allergen extracts in increasing doses. However, there is a risk in this method of allergic reactions or even anaphylactic shock. In order to minimise these risks, innovative preparations in the form of allergoids are employed. These are chemically modified allergen extracts which have significantly reduced IgE reactivity, but identical T-cell reactivity compared with the untreated extract (Fiebig, 1995, Allergo J. 4 (7): 377-382).

Even more substantial therapy optimisation would be possible with allergens prepared by recombinant methods. Defined cocktails of high-purity allergens prepared by recombinant methods, optionally matched to the individual sensitisation patterns of the patients, could replace extracts from natural allergen sources since these, in addition to the various allergens, contain a relatively large number of immunogenic, but non-allergenic secondary proteins.

Realistic perspectives which may result in reliable hyposensitisation with expression products are offered by specifically mutated recombinant allergens in which IgE epitopes are specifically deleted without impairing the T-cell epitopes which are essential for therapy (Schramm et al., 1999, J. Immunol. 162: 2406-2414).

A further possibility for therapeutic influencing of the disturbed TH cell equilibrium in allergy sufferers is immunotherapeutic DNA vaccination, which involves treatment with expressible DNA which encodes the relevant allergens. Initial experimental evidence of allergen-specific influencing of the immune response has been furnished in rodents by injection of allergen-encoding DNA (Hsu et al., 1996, Nature Medicine 2 (5): 540-544).

The present invention therefore also relates to a DNA molecule described above or below as medicament and to a corresponding recombinant expression vector as medicament.

The corresponding proteins prepared by recombinant methods can be employed for therapy and for *in vitro* and *in vivo* diagnosis of pollen allergies.
For preparation of the recombinant allergen, the cloned nucleic acid is ligated into an expression vector, and this construct is expressed in a suitable host organism. After biochemical purification, this recombinant allergen is available for detection of IgE antibodies by established methods.

The present invention therefore furthermore relates to a recombinant expression vector comprising a DNA molecule described above or below, functionally linked to an expression control sequence, and a host organism transformed with said DNA molecule or said expression vector.

The invention also relates to the use of at least one DNA molecule described above or at least one expression vector described above for the preparation of a medicament for the immunotherapeutic DNA vaccination of patients with allergies in the triggering of which group 4 allergens from the *Poaceae,* preferably *Triticeae*, in particular Sec c 4, Hor v 4, Tri a 4, are involved and/or for the prevention of such allergies.

As already stated, the invention can be used as an essential component in a recombinant allergen- or nucleic acid-containing preparation for specific immunotherapy. A number of possibilities arise here. On the one hand, the protein with an unchanged primary structure may be a constituent of the preparation. On the other hand, a hypoallergenic (allergoid) form can be used in accordance with the invention for therapy in order to avoid undesired side effects by specific deletion of IgE epitopes of the molecule as a whole or the production of individual fragments which encode T-cell epi-

topes. Finally, the nucleic acid per se, if ligated with a eukaryotic expression vector, gives a preparation which, when applied directly, modifies the allergic immune state in the therapeutic sense.

The present invention furthermore relates to the polypeptides encoded by one or more of the DNA molecules described above, preferably in their property as medicament.

These are proteins corresponding to an amino acid sequence in accordance with SEQ ID NO 2, 4, 6, 8 or 10. In particular, these are mature proteins (without signal sequence component), beginning with amino acid 23 (SEQ ID NO 2, 4 and 6) and with amino acid 22 (SEQ ID NO 8, 10). The invention furthermore relates to proteins which contain these amino acid sequences or parts of these sequences.

The invention accordingly also relates to a process for the preparation of such polypeptides by cultivation of a host organism and isolation of the corresponding polypeptide from the culture.

The invention likewise relates to the use of at least one polypeptide described above for the preparation of a medicament for the diagnosis and/or treatment of allergies in the triggering of which group 4 allergens from the *Poaceae*, preferably *Triticeae,* in particular Sec c 4, Hor v 4, Tri a 4, are involved and for the prevention of such allergies.

When determining the protein and DNA sequences according to the invention, the following procedure was followed:

<u>Sec c 4 from rye</u>

1. Starting from the DNA sequence of Phl p 4 (SEQ ID NO 12, WO 04/000881), specific primers (Table 1) derived from the Phl p 4 sequence

were generated. Five clones were obtained from rye pollen DNA by PCR with primers #87 and #83. The amplified Sec c 4 gene fragment 1 corresponding to these clones encodes a polypeptide corresponding to amino acids 68-401 of Phl p 4 (SEQ ID NO 12).

2. An EST database search was carried out with the partial Sec c 4 sequence. However, no homologous sequences were found in EST databases specialising in rye. Instead, individual, homologous, non-overlapping EST fragments were found in EST databases specialising in barley and wheat. Individual EST fragments extend into the 5'-UTR region and others into the 3'-UTR region (UTR = untranslated region) of the corresponding genes.

3. However, a complete group 4 gene from wheat or barley cannot be constructed from the EST sequences found in the databases since these sequences do not overlap and a homologous group 4 gene is not known. However, it was possible to assign these EST sequences with reference to the Phl p 4 sequence (SEQ ID NO 11) and the Sec c 4 fragment obtained in step 1 and these served as template for the preparation of PCR primers.

4. With the aid of primers #195 and #189 prepared in this way, three clones were obtained by PCR. Primer #195 was derived from a barley EST sequence, primer #189 is a Phl p 4-specific primer and overlaps the Phl p 4 stop codon and the codons of the 10 C-terminal Phl p 4 amino acids. The Sec c 4 gene fragment 2 amplified in this way encodes a polypeptide, beginning within the signal sequence and ending with the position corresponding to position 490 of Phl p 4. This polypeptide covers the N terminal of Sec c 4.

5a. Three further clones were obtained by PCR with primers #195 and #202. Both primers were derived from barley EST sequences. The ampli-

fied Sec c 4 gene 3 encodes the corresponding amino acids beginning within the signal sequence and ending at the C terminal of Sec c 4. The complete sequence of mature Sec c 4 is thus present in the sequence determined.

The next two steps 5b and 5c serve to double-check the result obtained in step 5a:

5b. A further clone was obtained by PCR with primers #195 and #203. Primer #195 was derived from a barley EST sequence, primer #203 from a wheat EST sequence. The amplified Sec c 4 gene encodes the corresponding amino acids beginning within the signal sequence and ending at the C terminal of Sec c 4. The complete sequence of mature Sec c 4 is therefore present in the sequence determined.

5c. A further clone was obtained by PCR with primers #195 and #198. Also primer #198 The amplified Sec c 4 gene encodes the corresponding amino acids beginning within the signal sequence and ending at the C terminal of Sec c 4. The complete sequence of mature Sec c 4 is therefore present in the sequence determined.

Two isoforms Sec c 4.01 and 4.02 were found. The mature allergens begin with amino acid 23 of the sequences in accordance with SEQ ID NO 2, 4 and 6.

<u>Hor v 4 from barley</u>

With the aid of the Sec c 4 sequences obtained as described above, homologous EST fragments were found in EST databases of *Hordeum vulgare*. These fragments overlap, but not to give a complete gene. With reference to the EST sequences found, however, it was possible to generate Hor v

4-specific primers, which were used for amplification of the Hor v 4 gene from genomic DNA.

In total, 15 clones were analysed.
4 clones were obtained by PCR with primers #195 and #198.
4 clones were obtained by PCR with primers #195 and #202.
3 clones were obtained by PCR with primers #194 and #198.
4 clones were obtained by PCR with primers #194 and #202.

The derived protein sequence begins within the signal sequence of Hor v 4 and extends to the C-terminal end of the protein (from amino acid 23 of SEQ ID NO 6).

Tri a 4 from wheat

With the aid of the Sec c 4 sequences obtained as described above, homologous EST fragments were found in EST databases of *Triticum aestivum.* These fragments overlap, but not to give a complete gene. With reference to the EST sequences found, however, it was possible to generate the Tri a 4-specific primers #199, #203, #204 and #206, which were used for amplification of the Tri a 4 gene from genomic DNA.

In total, 13 clones were analysed.
4 clones were obtained by PCR with primers #204 and #203.
4 clones were obtained by PCR with primers #204 and #199.
3 clones were obtained by PCR with primers #206 and #203.
4 clones were obtained by PCR with primers #206 and #199.

The derived protein sequences begin within the signal sequence of Tri a 4 and extend to the C-terminal end of the protein.
Two variants Tri a 4.01 (from amino acid 22 of SEQ ID NO 8) and Tri a 4.02 (from amino acid 22 of SEQ ID NO 10) were found.

In order to prepare the recombinant allergens according to the invention, the DNA sequences in accordance with SEQ ID NO 1, 3, 5, 7 and 9 were incorporated into expression vectors (for example pProEx, pSE 380). *E. coli*-optimised codons were used for the N-terminal amino acids known from the protein sequencing.

After transformation in *E. coli*, expression and purification of the recombinant allergens according to the invention by various separation techniques, the proteins obtained were subjected to a refolding process.
Both allergens can be employed for highly specific diagnosis of grass pollen allergies. This diagnosis can be carried out *in vitro* by detection of specific antibodies (IgE, IgG1 - 4, IgA) and reaction with IgE-loaded effector cells (for example basophiles from blood) or *in vivo* by skin test reactions and provocation at the reaction organ.

The reaction of the allergens according to the invention with T-lymphocytes from grass pollen allergy sufferers can be detected by allergen-specific stimulation of the T-lymphocytes for proliferation and cytokine synthesis both with T-cells in freshly prepared blood lymphocytes and also on established nSec c 4, nHor v 4 or nTri a 4-reactive T-cell lines and clones.

The triplets encoding the cysteines were modified by site-specific mutagenesis in such a way that they encode other amino acids, preferably serine. Both variants in which individual cysteines have been replaced and those in which various combinations of 2 cysteine residues or all cysteines have been modified were prepared. The expressed proteins of these cysteine point mutants have greatly reduced or zero reactivity with IgE antibodies from allergy sufferers, but react with the T-lymphocytes from these patients.

The present invention therefore furthermore relates to a DNA molecule described above or below in which one, a plurality of or all the cysteine residues of the corresponding polypeptide have been replaced with another amino acid by site-specific mutagenesis.

The immunomodulatory activity of hypoallergenic fragments which correspond to polypeptides having T-cell epitopes and that of the hypoallergenic point mutants (for example cysteine replacements) can be detected by their reaction with T-cells from grass pollen allergy sufferers.

Such hypoallergenic fragments or point mutants of the cysteines can be employed as preparations for hyposensitisation of allergy sufferers since they react with the T-cells with equal effectiveness, but result in reduced IgE-mediated side effects owing to the reduced or entirely absent IgE reactivity.

If the nucleic acids encoding the hypoallergenic allergen variants according to the invention or the unmodified DNA molecules according to the invention are ligated with a human expression vector, these constructs can likewise be used as preparations for immunotherapy (DNA vaccination).

Finally, the present invention relates to pharmaceutical compositions comprising at least one DNA molecule described above or at least one expression vector described above and optionally further active ingredients and/or adjuvants for the immunotherapeutic DNA vaccination of patients with allergies in the triggering of which group 4 allergens from the *Poaceae,* preferably *Triticeae*, in particular Sec c 4, Hor v 4, Tri a 4, are involved and/or for the prevention of such allergies.

A further group of pharmaceutical compositions according to the invention comprises at least one polypeptide described above instead of the DNA and is suitable for the diagnosis and/or treatment of said allergies.

Pharmaceutical compositions in the sense of the present invention comprise, as active ingredients, a polypeptide according to the invention or an expression vector and/or respective pharmaceutically usable derivatives thereof, including mixtures thereof in all ratios. The active ingredients according to the invention can be brought into a suitable dosage form here together with at least one solid, liquid and/or semi-liquid excipient or adjuvant and optionally in combination with one or more further active ingredients.

Particularly suitable adjuvants are immunostimulatory DNA or oligonucleotides having CpG motives.

These compositions can be used as therapeutic agents or diagnostic agents in human or veterinary medicine. Suitable excipients are organic or inorganic substances which are suitable for parenteral administration and do not adversely affect the action of the active ingredient according to the invention. Suitable for parenteral use are, in particular, solutions, preferably oil-based or aqueous solutions, furthermore suspensions, emulsions or implants. The active ingredient according to the invention may also be lyophilised and the resultant lyophilisates used, for example, for the preparation of injection preparations. The compositions indicated may be sterilised and/or comprise adjuvants, such as preservatives, stabilisers and/or wetting agents, emulsifiers, salts for modifying the osmotic pressure, buffer substances and/or a plurality of further active ingredients.

Furthermore, sustained-release preparations can be obtained by corresponding formulation of the active ingredient according to the invention – for example by adsorption on aluminium hydroxide.

The invention thus also serves for improving *in vitro* diagnosis as part of allergen component-triggering identification of the patient-specific sensitisation spectrum. The invention likewise serves for the preparation of sig-

nificantly improved preparations for the specific immunotherapy of grass pollen allergies.

Table 1 Primers used

a) Sec c 4

Primer number	SEQ ID NO	Sequence
#0083	30	GGCTCCCGGGGCGAACCAGTAG
#0087	31	ACCAACGCCTCCCACATCCAGTC
#0189	32	GATAAGCTTCTCGAGTGATTAGTACTTTTTGAT CAGCGGCGGGATGCTC
#0195	33	GCTCTCGATCGGCTACAATGGCG
#0198	34	CACGCACTACAAATCTCCATGCAAG
#0202	35	CATGCTTGATCCTTATTCTACTAGTTGGGC
#0203	36	TACGCACGATCCTTATTCTACTAGTTGGGC

a) Hor v 4

Primer number	SEQ ID NO	Sequence
#0194	37	GCCTTGTCCTGCCACCACGCCGCCGCCACC
#0195	38	GCTCTCGATCGGCTACAATGGCG
#0198	39	CACGCACTACAAATCTCCATGCAAG
#0202	40	CATGCTTGATCCTTATTCTACTAGTTGGGC

c) Tri a 4

Primer number	SEQ ID NO	Sequence
#0199	41	CACGCACTAAATCTCCATGCAAG
#0203	42	TACGCACGATCCTTATTCTACTAGTTGGGC
#0204	43	AAGCTCTATCGCCTACAATGGCG
#0206	44	GGTGCTCCTCTTCTGCGCCTTGTCC